

PRUEBA DATA SCIENCE

Instrucciones:

- La prueba consiste en responder una serie de preguntas a partir del conjunto de datos proporcionado junto con este instructivo.
- El candidato contará con 72hs. para realizar la prueba y enviar los resultados.
- El candidato podrá usar *Python* o *R*.
- Se puede consultar cualquier fuente de información que se considere oportuna (internet, libros, foros, etc.).
- La prueba es individual.
- El formato de la entrega podrá ser un script (.py o .R) o un notebook (.ipynb o .Rmd).
- Es importante justificar todas las decisiones que se tomaron.
- No se debe asumir en ningún caso que los datos no contienen errores.

Problema

Se dispone de información de venta de productos de la categoría Aguas Saborizadas para los meses de junio a noviembre de 2020 (formato en Anexo I).

NOTA: Los datos proporcionados son ficticios y han sido creados con el único fin de ser usados para esta prueba.

A partir de los datos proporcionados responder a las siguientes preguntas.

Pregunta 1

Obtener la lista de productos que se venden en al menos el 80% de los puntos de venta.

Pregunta 2

Ordenar los productos de mayor a menor según su venta en volumen medido litros (suma de los 6 meses) y quedarse con los que acumulan el 80% (Pareto).

Pregunta 3

Obtener para cada producto el identificador (pdv_codigo) de uno cualquiera de los puntos de venta que lo vende con mayor frecuencia.

Considerar solamente los productos que están presentes en al menos el 80% de los puntos de venta (*pregunta 1*) y en el top 80% de la venta en volumen (*pregunta 2*).

Para el cálculo de la frecuencia considerar solamente los días que el punto de venta estuvo abierto (tuvo alguna venta). Por ejemplo, si un producto en un punto de venta se vende todos los días de lunes a sábado y el punto de venta está cerrado los domingos, la frecuencia tiene que ser cada 1 día.

Pregunta 4

Calcula la variación porcentual de la venta de la categoría Aguas Saborizadas durante los meses de setiembre a noviembre, y comparalos con los meses de junio a agosto.

Pregunta 5

Identificar a partir de los datos proporcionados al menos una causa probable por la cual la venta de **SALUS FRUTTE CERO ANANA 1,65L** ha crecido en el mes de setiembre.

Pregunta 6

Utilizando los datos proporcionados de productos y ventas, desarrolle un modelo de forecasting para predecir las ventas futuras de los productos, por código de barras y punto de venta (pdv_codigo), teniendo en cuenta patrones históricos. Evalúe el rendimiento del modelo utilizando las métricas que considere más adecuadas. Además, comente cualquier inconveniente encontrado durante el desarrollo del modelo y proponga posibles soluciones para abordarlos.

ANEXO I - DATOS PROPORCIONADOS

Se proporcionan dos archivos: *productos.csv* y *ventas.csv* correspondientes a los productos y las ventas respectivamente.

Archivo productos.csv

Cada línea del archivo representa un producto.

Columna	Descripción	Ejemplo
codigo_barras	Es el código de barras impreso en el empaque del producto.	7734284001913
descripcion	Descripción del producto.	AGUA SABORIZADA H2OH! CITRUS 1,5L
marca	Marca del producto.	H2OH!
proveedor	Proveedor del producto.	FNC S.A.
contenido	Contenido en mililitros del producto.	1500

Archivo ventas.csv

Cada línea del archivo representa el acumulado de todas las ventas de un producto en una fecha para un punto de venta.

Columna	Descripción	Ejemplo
fecha_comercial	Fecha en la que se hicieron las ventas.	2020-03-01
pdv_codigo	Id del punto de venta (supermercado) en el que se hicieron las ventas.	10
codigo_barras	Código de barras del producto vendido.	7734284001913

imp_vta	Suma del importe de todas las ventas del producto en el punto de venta y en la fecha.	160
cant_vta	Suma de las unidades de todas las ventas del producto en el punto de venta y en la fecha.	2

El ejemplo anterior indica que en el punto de venta con id 10, el día 1 de marzo de 2020, se vendieron dos unidades de 7734284001913 (H2OH! CITRUS 1,5L). En este caso el precio unitario es de \$80.